BLOOD CANCER DIAGNOSIS USING DEEP LEARNING: ENHANCING ACCURACY IN LEUKEMIA DETECTION

BY

SABBIRA SULTANA

ID:211-15-14575

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

MR. SHAH MD TANVIR SIDDIQUEE
ASSISTANT PROFESSOR
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANURAY 2024

APPROVAL

This Project titled **"Blood Cancer Diagnosis Using Deep Learning: Enhancing Accuracy in Leukemia Detection."**, submitted by **Sabbira Sultana, ID :211-15-14575** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 12 January, 2025.

BOARD OF EXAMINERS

Dr. Fizar Ahmed (FZA) — **Chairman**
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mr. Abdus Sattar (AS) — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Ms. Zahura Zaman (ZZ) — **Internal Examiner**
Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Ahmed Wasif Reza (DWR) — **External Examiner**
Professor
Department of Computer Science and Engineering
East West University

DECLARATION

We so certify that we, **Mr. Shah Md. Tanvir Siddiquee, Assistant Professor, Department of Computer Science and Engineering,** Faculty of Science and Information Technology, Daffodil International University, have supervised this study. We further declare that no portion of my study, nor any portion of it, has been submitted for consideration for a degree elsewhere.

Supervised by:

Tsiddiquee

Mr. Shah Md. Tanvir Siddiquee
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Sabbira Sultana

Sabbira Sultana
ID:211-15-14575
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

Without the assistance and participation of several people during the last two semesters, this task would not have been feasible. To everyone who has helped us in any manner, we are incredibly appreciative.

First and foremost, we would want to sincerely thank God for His divine gift, which enabled us to successfully finish the Final Year Design Project (FYDP).

We express our sincere gratitude and debt of gratitude to **Mr. Shah Md. Tanvir Siddiquee, Assistant Professor** in the Department of Computer Science and Engineering at Daffodil International University in Dhaka, Bangladesh. Our supervisor has extensive expertise in Deep Learning and is very interested in using it to complete this job. This project would not have been feasible without his unending patience, academic help, unceasing encouragement, vigorous and frequent monitoring, constructive criticism, insightful advise, reviewing several subpar versions, and fixing them at every turn.

In addition to other academic members and personnel of Daffodil International University's Department of Computer Science and Engineering, we would like to sincerely thank the Head of the Department for his helpful assistance in completing our research.

We would also like to express our gratitude to all of our Daffodil International University classmates who participated in this conversation throughout the course of their studies.

Lastly, we must respectfully thank our parents for their unwavering patience and support.

ABSTRACT

Blood problems are frequently discovered by visual inspection under a microscope. It might help classify a number of blood-related disorders in order to facilitate the identification of different leukemia conditions. Leukemia is the most common cancer that affects the body's red blood cells. Even in the absence of any outward signs, it can progressively damage all of the body's internal organs, leading to a host of other illnesses. Early blood disease identification is difficult since current technologies take longer. The creation of a technique that might aid in the classification of leukemia prediction is described in the article. The accurate identification of blood cancer is the primary objective of this study. Certain leukemia diseases, such acute myeloid leukemia (AML) and lymphoblastic leukemia (ALL), prevent cells from growing and protecting every component of the blood, which can cause cancer and other illnesses that can harm the blood in various ways. This study examined three distinct blood cell classifications: neutrophils, a subset of normal blood cells, and two distinct kinds of cancer cells. Clinical methods are not particularly good in predicting leukemia because symptoms include fatigue, sickness, fatigue, and loss of appetite. In order to obtain the best results for identifying blood cancer utilizing the most precise categories possible using a deep learning approach, the strategy was designed to predict blood cancer using DL techniques for picture automated identification and splitting methodology. used 6000 photos in this study. In the process, we showcased and compared several deep learning models, such as Inception V3, Mobile Net V2, InceptionResnetv2, and VGG16 & VGG19. The assessments that were done on the all five models showed that get the best accuracy of 100% on the dataset, which is excellent. Following accuracy in the detection of a certain kind of blood cancer.

Keywords: Acute Myeloid Leukemia (AML), Lymphoblastic Leukemia (ALL), Neutrophils Inception V3, MobileNet V2, InceptionResNetV2, VGG16, VGG19.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

Deadly disease cancer that typically results from a confluence of several hereditary and genetic issues that worsen over time. Malignant cells can appear as deadly, aberrant growths in other parts of an individual's body. While each technique has unique issues, several common causes of mortality are difficult histories, subpar evaluations, and subpar treatments. Investigating and testing machine learning techniques for leukemia diagnosis is the goal of. Acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL), and chronic myeloid leukemia (CML) are four of the main forms of malignancy. Blood is made up of three main types of cells: platelets, which are white blood cells that help form a clot, red blood cells, which are known to carry oxygen, and white blood cells called white blood cells that fight infections. The cells known as make up 93% of the millions of new blood cells the human bone marrow creates every day. The immune system produces more white blood cells than is really needed when a person has cancer, and both these types of leukemic cells are unable to fight the disease in the body in the same manner that healthy white blood cells do. As a result, over time, both external and internal organs start to work differently. As soon as a rare kind of leukemia is identified, treatment must begin right away. Certain bodily organs produce an excessive number of blood cells when leukemia is still present. Some forms of chronic leukemia don't seem to show any symptoms and are left untreated for a long time. Medical technology can identify leukemia by looking at neural lines and bone marrow samples.

This paper investigates blood cancer. Studies on acute myeloid and acute lymphoblastic leukemia are carried out. These medical evaluations are more expensive and painful, and it could take longer to receive the exact results. Thus, using machine learning techniques speeds up and simplifies the screening process. The machine learning system will receive its data input from the collected blood smear photographs. The final accuracy result of the model after segmentation and extraction.

Using image assessment can support the following specific objectives:

- To get the accuracy of three class predictions from five models.
- Using DL models to measure the three kinds of cells.

The related RGB (Red, Green, and Blue) views are transformed into grayscale versions in the image preprocessing system once the cancer cell photos have been read and acquired. Contrasted manipulation is one of the tone-enhancing techniques that is used to improve the amount of contrast. After the first processing is complete, it has been proposed that all of the images be segmented. This will lessen the complexity of the image and allow each component to be further examined or altered. Not to mention, there are picture classification techniques that may be used to recognize images of cancer. A range of deep learning models are available for picture classification, such as Inception V3, Mobile Net V2, InceptionResnetv2, and VGG16 & VGG19. Assessing the precision of blood cancer forecasts across all classes has been made easier with the use of these categories. Furthermore, a DL model for the identification of cancerous cells in the bloodstream was recently constructed using the three different forms of data that make up the DL model.

1.2 Motivation

I found making goods for the medical industry to be fascinating because it is the most crucial element of any healthcare websites. According to the system and my own, the medical sector has a major influence on all disease categories, including blood cancer. We are driven to use AI, ML, and DL in the health industry in order to help patients with blood cancer because of this. After considerable consideration, I was unable to come up with a paper topic that would satisfy the study's requirements. I thus sought advice from one of my esteemed teachers. She was pleased that I had chosen medicine as my area of study and recommended that I research a subject related to this field because blood cancer is so common in the present world. For this reason, I choose the topic " **Blood Cancer Diagnosis Using Deep Learning: Enhancing Accuracy in**

Leukemia Detection." Furthermore, I see that modern discoveries are being applied by society to improve the medical field, and that academics are increasingly studying the medical professions in an attempt to advance our own. We were inspired to carry out this type of research-based activity by the following. The world we live in depends on deep learning and machine learning techniques, which are all connected by artificial intelligence.

1.3 Objectives

We live in an era of rapid technical advancement. Any difficulty may be solved by innovation. Consequently, the expansion of the medical industry has been aided by a significant amount of scholarship. The largest threat to the health care industry throughout this period has been shown to be cancer. Human bodies can develop a variety of cancer forms. The primary goal of this work was to use deep learning methods to detect blood cancer cells. The primary objective is to properly classify various cancer cells in order to predict the cancer cells. There are other types of blood cancer, such as acute myeloid leukemia (AML), acute lymphoblastic leukemia (ALL), and neutrophils, which are healthy blood cells. My goal is to use an image processing method to classify blood cancer using dl. I may thus outline the goals as follows:

- To assist patients and the healthcare sector.
- To forecast the three different kinds of blood cancer photos using deep learning.
- Creating a system that recognizes the three forms of blood cancer using a deep learning model.
- To gather information that will aid in cancer prediction.
- Gaining a thorough comprehension of the deep learning domains.
- Using a variety of approaches to improve results.

1.4 Methodology

I want to provide a thorough explanation of the methodology and procedures used in my study's classification of blood cancers in this section of the paper. Some essential elements include the compilation of images, analysis, and suggested model, which is then further displayed using a suitable surface, diagram, formula, and summary. Applying that study to a dataset in my accomplishments involved merging classification models with deep learning forecasting and demarcation, which offer the highest efficiency and forecast the proper category of cancer. An overview of the statistical paradigms I used in my study and a list of the explicit requirements for implementation round out the chapter. In order to create this model, we used several deep learning computations. For this investigation, we used Deep CNN (Mobilenetv2, InceptionResnetv2, Inception V3, VGG19, VGG16). Computations that the model uses to classify data. I used three different group classifications for the models in this investigation. While most blood cells are malignant, I focused on two disease groups in this study: acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML), as well as neutrophils, a class of healthy blood cells. Every model in the study I carried out was trained using a variety of classes on every picture.

1.5 Project Outcome

The application of a basic principle to the classification of blood malignancies. A variety of online data from the internet is first merged with other resources to build a dataset. After that, data was added, the image was scaled, and labeled. This data might then be made available to the machine. Use data to refine and test learning procedure on new datasets. Consequently, we will be able to achieve our objective of having the algorithm predict blood cancer cells for each of the three types of blood cancer with the highest accuracy of all DL models. The primary challenge of this endeavor was assembling and analyzing the picture data from these datasets since managing them was challenging. I used a variety of methods and instruments to standardize and clean the dataset. We were left waiting for a very long time to view the finished output because the datasets were quite large and had many layers with different epochs. More datasets were available in this field. Since I didn't perform any research, I have to work really hard on any activity involving research in order to get the best answers and complete it as soon as possible. Again, there were problems with the preparation of picture data for classification when using the CNN model.

- Collecting data.
- Preparing data.
- Achieving 90% accuracy at the greatest conceivable level.
- Choosing every deep learning model based on test results.

1.6 Organization of the Report

In Chapter 1, the study's objectives, worries, research questions, and anticipated results were described. The report's overall structure is also covered in this section.

Chapter 2 contains all previous studies conducted in this area. In the section that follows, they give an illustration of the breadth that arises from their limitation of this study topic. The main challenges or obstacles to this inquiry were the subject of the last conversation. This chapter includes parts on relevant research summaries and studies, as well as a discussion of the challenges faced throughout the project's inception.

Chapter 3 offers a theoretical evaluation of the study's findings. This chapter contains further information, especially those arithmetic portion. This section also includes examples of how deep learning techniques are applied in the real world. The techniques for gathering data and the structure for putting it together are covered in the next section. In the last step, the model is assessed using a single-family confusion assessment matrix, which also yields a suitable tag for classifier identification. The study subject and methodology, operational effectiveness, data collecting strategy, data processing, recommended methodology, methodology of instruction, and prerequisites that must be satisfied for the research project to move forward are all included in this part. This study provides a comprehensive rationale for each model and DL approach employed.

The performance evaluation, results discussion, and experimental findings are presented in Chapter 4. To help with the project's implementation, a few test photographs are included in this section. A summary of the uses of supervised learning techniques closes this section.

Chapters 5 and 6 offered a summary about the study, Engineering Standards and Design Challenges, compliance with standards, financial analysis, engineering problem, information on the upcoming events, and a question regarding the results. This chapter offers an authorized sample to show whether or not the supplied report satisfies all requirements. Impacts on the surroundings and the group as a whole.

CHAPTER 2

Background

2.1 Introduction

Typically, an overview of the research, associated functions, and research challenges are included in this section. In the section labeled "Associated Work," I'll go into research publications by other authors and talk about how our methodology and accuracy compare to theirs. I'll veracity, and substance of other research in the section on related works. I will give an overview of our linked research at the study review unit. I make the decision to go over every problem we ran into while doing this study and how we raised the accuracy layer in the problems part. Everything was discussed beforehand.

2.2 Literature Review

A study on CNNs and DL focusing on the identification of cells known as blast cells in acute myeloid leukemia at the cellular stage was given by Christian Matek et al. in 2019 [1]. Malignant white blood cells must be detected in order to provide an appropriate diagnosis of hematologic malignancies such as acute myeloid leukemia. Expert human examiners perform microscopic morphological analyses of blood cells. One hundred patients with various AML types underwent peripheral blood smear investigation at the Leukemia Diagnostics Laboratory at Munich University Hospital. Using this dataset, they trained and evaluated a state-of-the-art image recognition method called ResNeXt CNN.18 classes were used to classify 18,000 single-cell images. They use k-fold cross validation techniques to do five-fold cross-validation by randomly stratifying and splitting the cell images into five folds. The CNN method exhibits remarkable accuracy, achieving a ROC AUC of about 0.99 in both mature and adolescent leukocyte situations. They [1] developed the initial idea, selected the cohort, digitalized blood smears, made software for annotation, trained, and evaluated the network. They also contributed to the cohort selection process and annotated the picture data. wrote the paper after analyzing the information. Every author gave their approval to the paper.

According to Rahul Duggal et al.'s 2017 definition [2], pixel stain quantities in medical imaging offer a fundamental knowledge of the interactions between tissues and stain chemicals. Their deep CNN network architecture, which incorporates the stain deconvolution layer (SD-Layer) to include tissue and cell staining, is suitable for biomedical microscopic imaging. There are 8938 cell nuclei in the dataset overall, 4469 of which are in the mean 2 class. For detection, they recommended the T-CNN and Alex Net models. By utilizing the recommended SD-Layer prefixed with the T-CNN architecture, they attain a 5-fold cross-validation accuracy of 93.2%. This suggests that a more precise portrayal of the input picture is obtained when employing SD-Layer. During the conditioning stage, they employed 180-degree random rotations as well as vertical spins in every period as data enrichment techniques. Both of these were positioned within a 400 by 400 black-colored area. making one's own dataset using both classes.

A study titled Classification of Acute Lymphoblastic Leukemia from Microscopic Images Using Convolutional Neural Networks was published in 2019 by Jonas Prellberg et al. [3]. Acute lymphoblastic leukemia (ALL) is a blood cancer that develops when abnormal lymphoblast cells multiply until a leukemia cell mass that is lethal is eventually accumulated, changing white blood cell microscopic images into benign and normal B-lymphoid progenitors utilizing a simple yet effective procedure. The C-NMC dataset is used to optimize a new convolutional neural network design with promising classification performance. They offer an easy to use yet effective technique of categorization using a Res Net convolutional neural network (CNN). The C-NMC dataset has around 10,000 pictures that feature Squeeze-and-Excitation modules.The plan of action 88.91% was the cumulative F1 score obtained on the examination of the C-NMC virtual test.

Vaibhav Rupapara et al. used machine learning to predict blood malignancy by combining hybrid logistic vector trees with leukemia microarray gene data [4]. The use of data from a ml method and genetic testing for cancer diagnosis is currently. This article proposes a supervised machine learning approach for blood cancer sickness prediction. A similarity of ml models, such as the RF, LR, SVC, KNN, NB, ETC, DT, ADA, and proposed LV Trees, are used for classification in this work. In the current study, a 22,283 gene leukemia microarray gene dataset is employed. To deal with uneven and big data sets, ADASYN replication as Chi square (Chi2) characteristics choice procedures are utilized. In order to spread the historical data for each target class equally, ADASYN creates synthetic data. Chi2 then selects the top 22,283 features to train learning models on, and LVTrees outperform all other models when ADASYN and Chi2 approaches are used. The findings of K-fold cross-validation show that the recommended models with k=10, or 10 cross-validation, are superior. The author of the notion and analysis is [4]. carried out the data curation and formal analysis. supplied the programs and tools. overseen the conducted tests and finished the write-up, write-review, and editing. All the authors reviewed the paper.

A research titled "Using an Ensemble of Convolutional and Recurrent Neural Networks, Normal and Leukemic Blast Cells in B-ALL Cancer is Classified" was published in 2019 by Salman Shah et al. [5]. In this work, immature lymphoblast and normal cells are distinguished using a custom-built deep learning model. Recurrent and convolutional neural networks are used to create the model. Additionally, it combines an RNN with a discrete cosine transform to take use of the cells' spectral properties. They may achieve notable performance gains when comparing their approach to CNN (AlexNet), a sequential DCT-LSTM model, and a hybrid CNN-RNN (LSTM-DENSE). The whole dataset collection consists of 10,661 training pictures from 76 individuals. DenseNet121, AlexNet, VGG19, and VGG16 models have all been used. Three-fold cross validation was used since the dataset was initially divided into three folders based on variations in topic level. Their hybrid CNN-RNN (LSTM-DENSE) model accomplishes the highest accuracy of 86.6% when k-fold cross validation techniques are used.

This study attempts to provide a mechanism for the diagnosis of leukemia using image processing techniques by Akash Holker et al. [6] in order to automate the detection process. Leukemia (blood cancer) is characterized by the proliferation of many aberrant cells. Acute lymphoblastic (ALL), acute myeloid (AML), chronic lymphocytic (CLL), and chronic myeloid (CML) leukemia are the four most common types. A file containing 1000 images of both cancerous and non-cancerous samples is created and kept for the module's training. The machine learning module that has been taught will produce an output indicating if the malignant or non-cancerous diagnosis was made using CNN model training. Machine learning yields fast and accurate results, reducing the cost of diagnostics and facilitating early diagnosis.

Machine Learning Algorithm for Leukemia Detection was published in 2021 by K P Jayavikash et al. [7]. Blood issues are frequently discovered by visual inspection under a microscope. This may facilitate the classification of several blood-related diseases. The paper details the creation of a technique that might help in leukemia diagnosis and detection early on. In order to identify leukemia, picture capture, segmentation, feature extraction, and classification were employed in this work. The images are from the ALL-IDB digital archive. A total of 108 photos were collected for the collection in 2005. The output of the model will show whether or not cancer is present in the input image. The accuracy rate of the dataset that follows is 93%.

According to M. Iswarya et al. [8], machine learning for leukemia detection will be defined in 2022. The main objective of this study is to use peripheral blood smear image analysis to detect leukemia in its early stages so that focused and efficient therapy may be administered. They use a Kaggle dataset of 1000 sample photographs as input datasets for their research. They proposed a convolutional neural network method that builds a model for leukemia identification in the ANACONDA program. The accuracy of detection with the CNN model is demonstrated to be greater than 90%. Lastly, data classification using CNN is utilized to identify leukemia.

2.2.1 Related research

Table 2.1: previous work

SL	Paper	Methodology	Description	Outcome

01	[4]	Machine Learning, Deep Learning, Convolutional Neural Network, Fully Connected, Neurons.	Recognize therapeutic plants and how to use them.	When using images from open field land zones for testing, they demonstrated a remarkable 85% accuracy rate. .

02	[5]	Feature extraction, Classification, Convolutional Neural Network	Determine the Medicinal Plants Based on the Outward Features of the Leaves and Flowers	Upon attempting over ten plants, identification rates as high as 98% have been achieved.
03	[6]	Performances based on ExG-ExR, Logistic Regression, Machine Learning	Employed Machine Learning Methods for Instantaneous Identification of Therapeutic Plants	The plant species are identified by the surface highlights and darkening on each divided leaf using a 93.3% precision Calculated Relapse classifier.

04	[7]	Stochastic Gradient Descent (SGD), k Nearest Neighbor (kNN), Support Vector Machines based on Radial Basis Function Kernel (SVM-RBF), Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA) classifiers are used to examine the classification performance.	Using channels with the best color spaces and textures Recognizing the leaves of therapeutic plants	The results of characterizing a dataset including 250 leaf images featuring five distinct plant species demonstrate an identification rate of 96.7 percent.
05	[8]	Convolutional Neural Network(CNN),YOLOv2	With Yolo Neural Network Classification and Recognition of Bangladeshi Plant Leaves	Have been used CNN, YOLOv2.The accuracy rate was 96%.
06	[9]	Random forest classifier, k-nearest neighbor, naïve Bayes, support vector machines and neural networks.	Through the use of machine learning techniques Automatic Identification of Therapeutic Plants	The random forest classifier fared better than other Machine Learning techniques, with

				an exactness of 90.1%.
07	[10]	Employed Medicinal Plant Classification; Transfer Learning; Deep Learning; MobileNet.	Combining Transfer Learning and Deep Learning for the Classification of Medicinal Plants	They have made use of Deep Learning, Transfer Learning, Medicinal Plant Classification, and MobileNets. They also have a 98.7% accuracy rate.

2.3 Gap Analysis

Deep learning-based blood cancer detection has advanced significantly, but there are still precision, usefulness, and uptake of AI based treatments in clinical settings. These include addressing ethical issues, guaranteeing continuous performance, connecting deep learning technologies with clinical procedures, increasing model interpretability, and improving dataset quality. Patients with blood cancer may benefit from earlier, more precise diagnosis and improved outcomes thanks to current developments in deep learning models.

2.4 Summary

This section contains all previous studies conducted in this area. In the section that follows, they give an illustration of the breadth that arises from their limitation of this study topic. The main challenges or obstacles to this inquiry were the subject of the last conversation. This chapter includes parts on relevant research summaries and studies, as well as a discussion of the challenges faced throughout the project's inception.

CHAPTER 3

Research Methodology

3.1 Methodology/Requirement Analysis & Design Specification

3.1.1 Overview

I want to provide a thorough explanation of the methodology and procedures used in my study's classification of blood cancers in this section of the paper. Some essential elements include the compilation of images, analysis, and suggested model, which is then further displayed using a suitable surface, diagram, formula, and summary. Applying that study to a dataset in my accomplishments involved merging classification models with deep learning forecasting and demarcation, which offer the highest efficiency and forecast the proper category of cancer. An overview of the statistical paradigms I used in my study and a list of the explicit requirements for implementation round out the chapter. For this investigation, we used Deep CNN (Mobilenetv2, InceptionResnetv2, Inception V3, VGG19, VGG16). computations that the model uses to classify data. I used three different group classifications for the models in this investigation. While most blood cells are malignant, I focused on two disease groups in this study: acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML), as well as neutrophils, a class of healthy blood cells. Every model in the study I carried out was trained using a variety of classes on every picture.

3.1.2 Proposed Methodology

Various techniques or approaches might be utilized to determine how to assess the information utilized in this investigation. A multi-step methodology that included model creation, extension and improvement, data gathering, and production was applied in this work.

Step 1: Data Collection: Using Kaggle, I gathered and examined web statistics data to produce a reliable collection of my own. Due to the challenges in locating and getting data for the various blood cancer cases, pictures of the multiple blood images obtained at different stages, and other factors, a big, complete dataset is not easily.

Step 2: Data preparation: Each piece data treated independently after being gathered in its raw form many blood cancer providers. Many data sets can have errors, especially ones that contain noise. In technical terms, I use the chosen data set to go on to the next phase after processing the information first.

Step 3: Data preprocessing: The results expanded and narrowed as each class was assessed. I had to alter the size and add information for it to work. I made just the biggest and most socially acceptable changes since I was concerned about overfitting.

Step 4: Model Selection: Once you've made your decision, train and assess the selected model using the provided data to increase accuracy. DL uses a wide variety of models.

Using my technology, I experimented with many versions of the concept to determine the best configuration for precise data calculation.

Step 5: Performance Evaluation: An description of each outcome is given in this section. Even after testing and training, these strategies left us lacking adequate reliability for the next two courses. They created a technique for classifying different medicinal plants and created graphics for f1 measurement, recall, efficiency, and confusion matrix.

Step 6: Concluding Remarks and Future Projects: A timetable and an overview of the changes are provided in the section that follows.

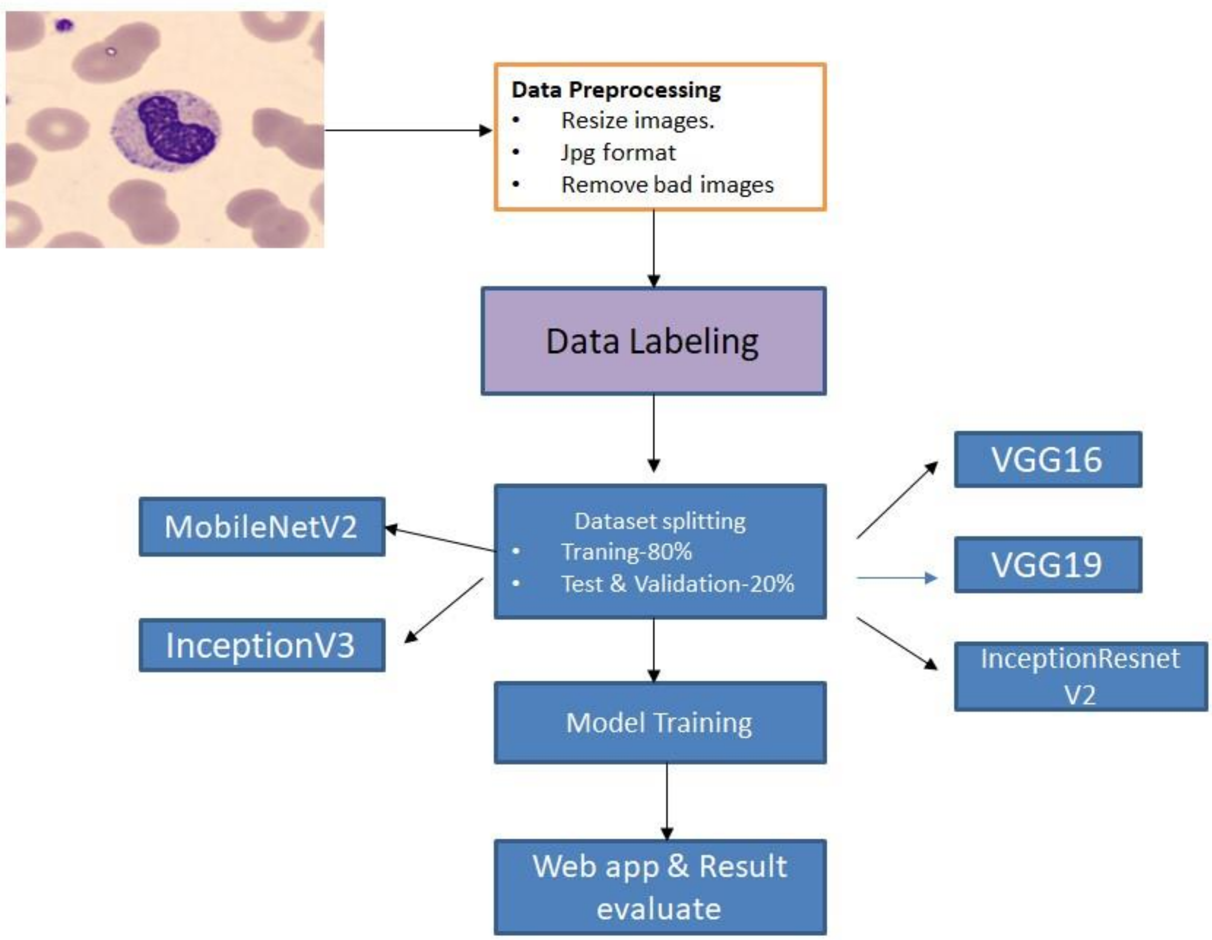


Fig 3.1: Entire Proposed model.

Figure 3.1 illustrates the application of a basic principle to the classification of blood malignancies. A variety of online data from the internet is first merged with other resources to build a dataset. After that, data was added, the image was scaled, and labeled. This data might then be made available to the machine. Consequently, we will be able to achieve our objective of having the algorithm predict blood cancer cells for each of the three types of blood cancer with the highest accuracy of all DL models.

3.1.3 Implementation Requirements

A research area for managing modeling, data collection, job completion, and providing instruction, in addition to execution. We provide the tools and techniques we use as measurement experts. We utilized the Windows operating system of the organization, NumPy, SkLearn, and OpenCV in addition to Python as our programming language. Google Colab was the platform of choice for all instruction and assessment. Python programmers may use Google Colab to build code for deep learning and data science techniques.

Libraries used:

- **Matplotlib**: The Pyplot suite of Matplotlib utilities facilitates scoring, organizing, graphing, and plotting. This might be used in form-building to highlight specific narrative stances or the limits of the story's credibility.
- **NumPy**: Vector processing has been made simpler by the language's NumPy package. This subject covers in-depth explanations of conversion indices, matrix computations, and the inverted Fourier transform. Many methods and tools for handling various types of matrices are included in the NumPy Python packages. NumPy can help make the process of designing gadgets more logical and beneficial. To put it briefly, NumPy is a Python package designed mostly for numerical analysis. "estimates. Py" is another option for representing this.
- **Sklearn:** A robust and user-friendly modeling and analysis of information tool. NumPy, SciPy, and Matplotlib are three Python tools that were utilized in the layout. These are freely used, publicly available, and open-source tools.

- **Seaborn:** The next version of this well-known Python data visualization program will use Matplotlib. This is an easy-to-use tool for creative data visualization.
- **H5.py:** The Python h5py package enables users to read HDF5 code that has been fragmented. A large portion of the data—mostly integers—is handled by NumPy, and HDF5 is used for archiving thereafter.
- **TensorFlow**: It appears that the general public may access this set of AI technology. It offers an extensive set of tools for creating and utilizing different machine learning theories, the most notable of which are neural network-based models. Tensor Flow's adaptability, effectiveness, and ease of use make it a viable choice for a variety of intelligence-based applications.
- **Pandas**: Pandas is a freeware toolbox for analyzing and manipulating data specific to a language. Basic data kinds and statistical analysis techniques can help ensure organized data administration, particularly when working with summary data.
- **OS:** A variety of features provided by the Py System element allow employees to converse with one other in the same language.

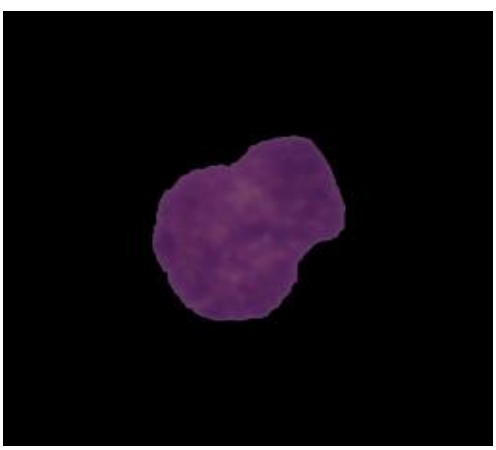
In order to facilitate in-depth research, the hardware, software, and data resources employed in this study's " **Blood Cancer Diagnosis Using Deep Learning: Enhancing Accuracy in Leukemia Detection.** " configuration have been carefully selected. Computing systems with a processing core composed of strong CPUs and powerful GPUs are referred to as high-performance computers. I understand that no invention can promise perfect results. Similar to that, that we may adjust our model's parameters as it's being trained to increase accuracy.

3.2 Detailed methodology & Design

The vast majority of the research I did focused on the many methods society provides. Five different approaches have been used to classify blood cancer and identify it using a deep learning network. For my particular dataset, I used a number of methods. In this case, a dataset I put together from several web sources served as the primary source of data. I will be able to assess the reliability of the seven methods I used and study factors like the impact of extra data I provided using the same source. The new data is an exact replica of the old dataset with which it was combined. Tags can be used to describe their meaning by grouping them into related categories and classes. This model was developed using a number of deep learning calculations. We employed Deep CNN (Mobilenetv2, InceptionResnetv2, Inception V3, VGG19, VGG16) for this study. Calculations that the model makes in order to categorize data. For the models in this study, I employed three distinct group classifications. Although the majority of blood cells are cancerous, I concentrated on two disease categories in my study: neutrophils, a subset of healthy blood cells, and acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML). In the study I conducted, each model was trained on each image using a range of classes.

3.2.1 Data Collection

The dataset was gathered via the internet resource Kaggle. I made a dataset of 6000 images. Based on the kind of sickness each class in the collection reflects, it is separated into three groups with 2000 photographs each. 3 types of blood sample like: Neutrophil-containing normal blood cells, acute myeloid leukemia, and acute lymphoblastic leukemia are among the classifications.



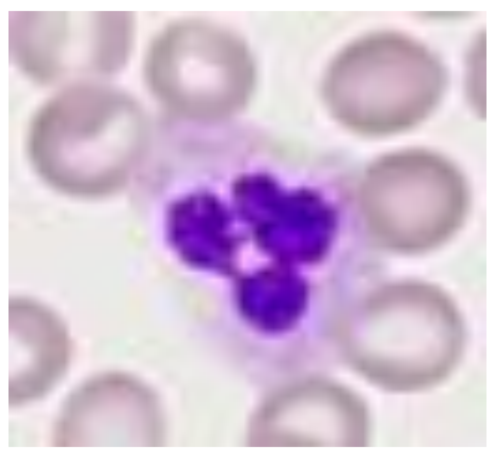


Fig 3.2: Acute Lymphoblastic Leukemia. **Fig 3.3: Acute Myeloid Leukemia.**

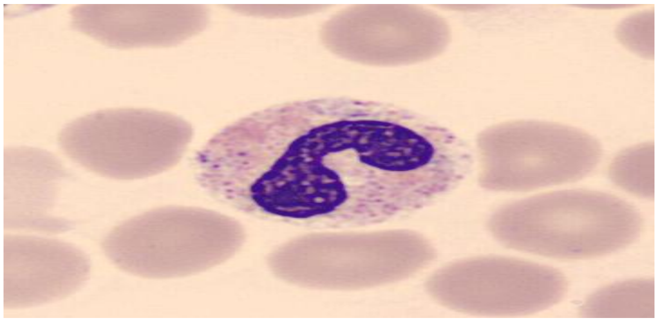


Fig 3.4: Normal Blood cell Neutrophils.

All of the data items found in each of these files are shown in Table 3.1, which organizes them into five main categories:

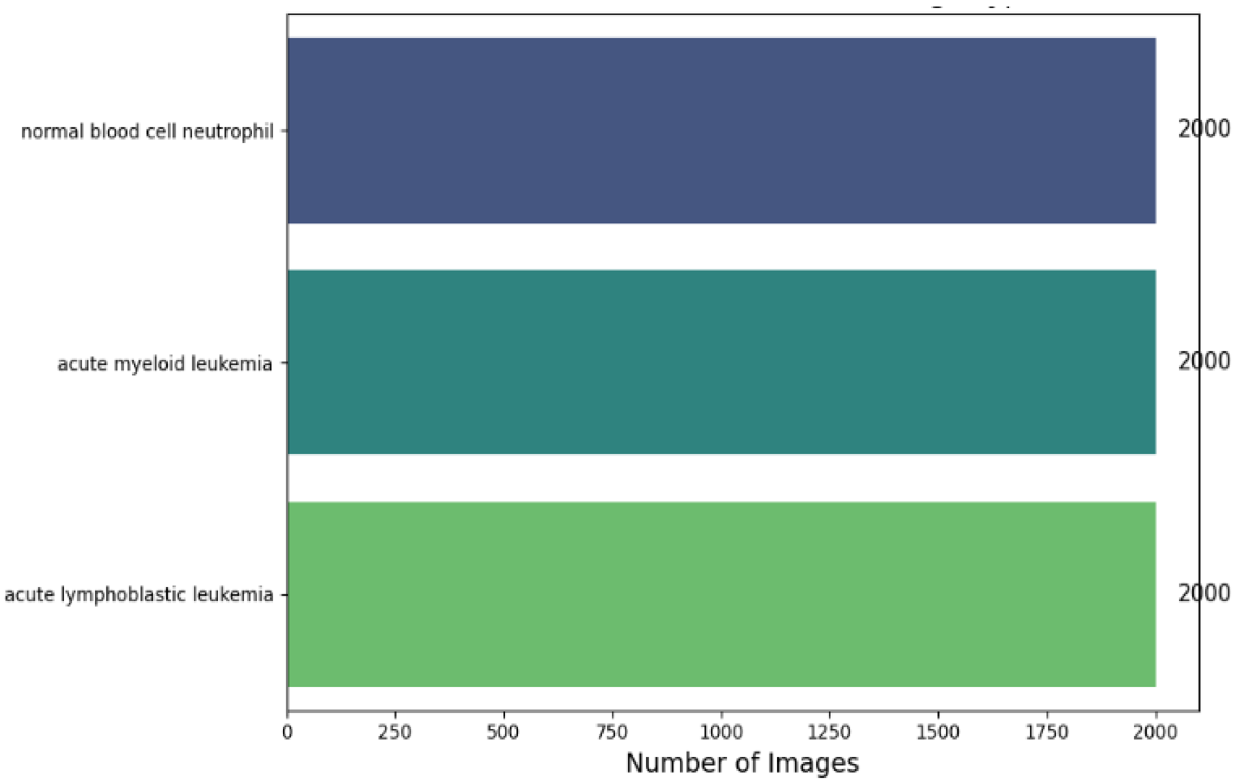


Fig 3.5: Three classes of data.

3.2.2 Data Preprocessing

The main element of a dataset used in data processing. This research has all the image data required to provide reliable findings. A data set's data processing methodology is crucial. Our ability to work with data is reliant on preprocessed data. For the many kinds of blood cancer, a few online resource data sets were integrated to create a comprehensive dataset of three types for this investigation. Pretreatment of the data is typically critical to the effectiveness of dataset modification. Better pre-processed results will yield more accurate results. To put it simply, the main barrier for this kind of based research task is this. Information preparation and data resizing are two of the two procedures in a system for the processing of data.

- **Data gathering and preprocessing:** Kaggle had collected various online field data, such as height and breadth, which were used to make each image in the collection I used. Since every image for my model must meet strict quality standards, I modified the script to produce an image that is 500 by 500 pixels. Additionally, I prefixed each with "jpg" before processing it. After data augmentation, I divided and edited the pictures to prepare them for classification. To train the framework, I thus used the split versions of each dataset.
- Images whose sizes are specified by codes.
- File types in JPG will now be utilized.
- Eliminate any inaccurate pictures.
- Deleted unnecessary images.

3.2.3 Data for Training, Testing and Validation

Deep learning is an emerging field that includes investigating and developing methods for extracting data from results based on that understanding. After reading the incoming data and generating an equation from it, these algorithms employ the equation to examine the data and make inferences. Before being utilized in the process of creating the model, in the process of constructing a model: train, valid, and test. Initially, 20% and 80% of the training dataset should be used for learning, and the remaining 50% should be used for testing and validation.

3.2.4 Applied Model

1. **InceptionV3:** CNN InceptionV3 serves as the basis for categorizing and identifying pictures. It's one of the ideas that the staff of the firm came up with for the group of Inception DL thoughts. Due to its innovative of Inception: Volume Three. Levels with optimal pooling, additional classifiers, convolutional layers, and fully linked layers are examples of combinations. Networking's multilevel randomization and periodic mixing of the Creativity modules allow it to gather data at several levels of abstraction.

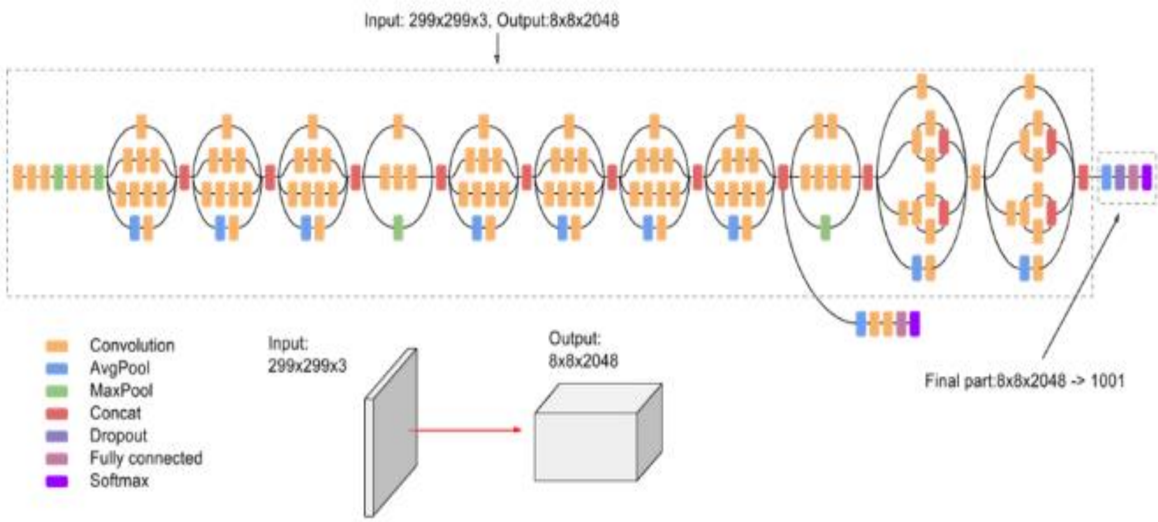


Fig 3.6: The architectural model of Inception V3.

2. **InceptionResNetV2:** First, two additional deep neural network designs, ResNet and Inception, are combined to form ResNetV2, a neural network architecture. With residual connections, it is an extension of the ResNet-essential Inception architecture. ResNetV2, which Google researchers claimed to be a network from the Inception group, made its debut. It is renowned for demonstrating remarkable proficiency in a range of artificial intelligence applications, like as object and picture identification. It demonstrated state-of-

the-art capabilities when it was used in the ImageNet massive numbers visual recognition challenge (ILSVRC).

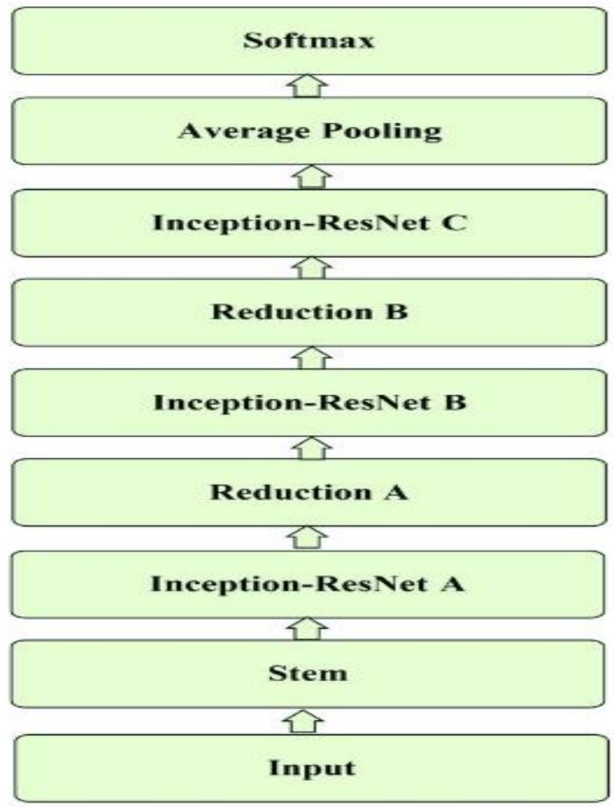


Fig 3.7: The architectural model of InceptionResNetV2.

3. **VGG16:** VGG16, the abbreviation for this model, is sometimes referred to as VGG Net. The method makes use of neural networks with sixteen-layer convolutions (CNNs). Considering that pretrained neural networks There are more than a million photos in ImageNet. A mouse, pencil, internet keyboard, and other items are among the 1000 unique item categories that the conditioned prior to use network can identify in images. To train the network, may be able to handle the 224 × 224 image source size. For other pretrained network possibilities, view the Originally trained Parallel neural network models in MATLAB.

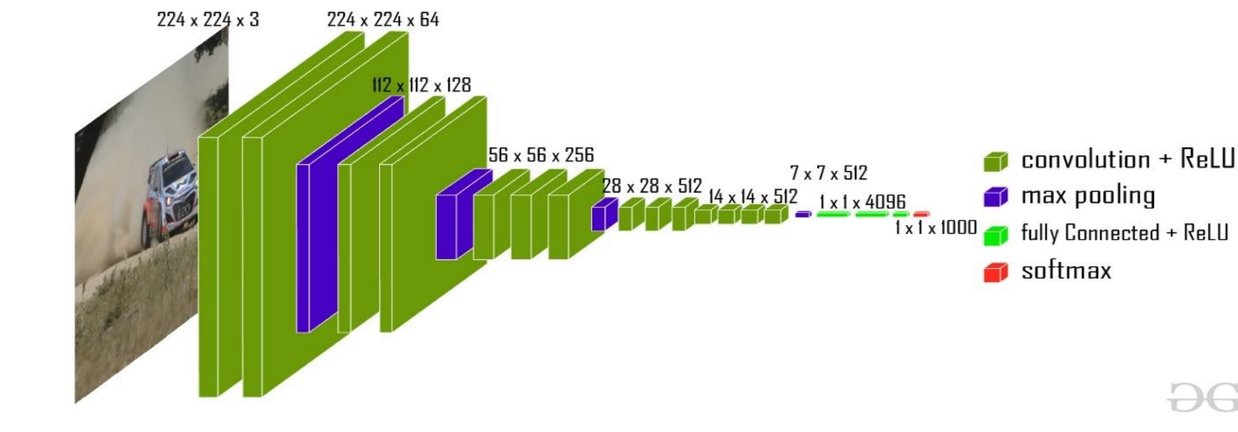


Fig 3.8: The architectural model of VGG16.

The picture size that is transmitted to the network is (224, 224, 3). The first two layers have a 3*3 filter area and sixty-four channels with the same padding. The max pool layer (stride 2, 2) will be followed by a second layer including convolution layers with filter sizes (3, 3) and 128 filter measures. The largest pooling layer, which is the same as the one before it and has the highest frequency (2, 2), comes next in sequence. A set of 256 filters with filtered widths of three and triple comes after two convolutional layers. Convolution is used to create two sets of three independent layers, followed by a max pool layer. Every one of the 512 filters has the same dimensions and spacing (3, 3).

4. **VGG19:** A variant of the VGG model is the VGG19 model. Nineteen convolutional layers, three fully connected layers, five maximum pooling layers, and one SoftMax layer make up its composition. The elements of the matrix were 224, 224, was sent to the system. The one preprocessing step of the system was to take away each pixel's mean RGB value, which was calculated for the whole training set. They used spatial padding to keep the image's depth. Speed 2 above two * two-megapixel panes was needed to get maximal pooling. Next, the model classification was enhanced, computation speed was increased, and non-linearity was increased by utilizing the modified uniform constituent (ReLu). When compared to earlier models that used tanh

or exponential functions, the ReLu proved to be significantly superior. Three of the 4096-sized layers that were produced were completely connected. The last layer is referred to as the softest max functions, and for a 1000-way classification, an extra layer with an accompanying channel with a frequency of 1000 is added.

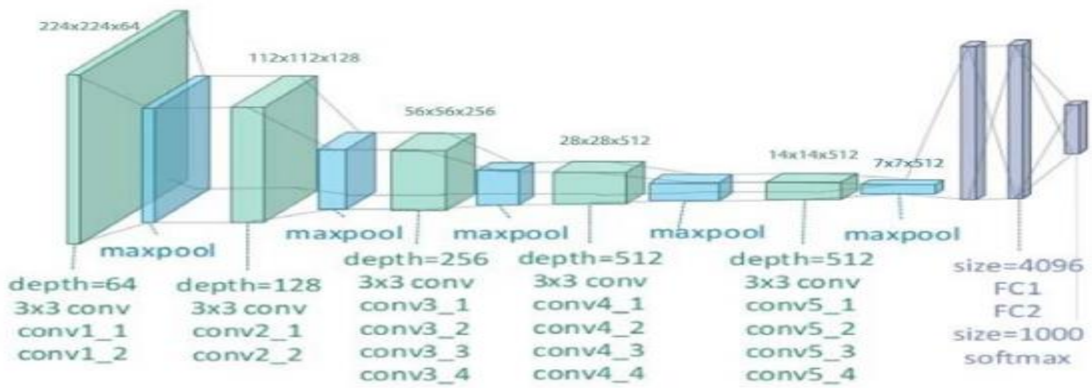


Fig 3.9: The architectural model of VGG19.

5. MobileV2: MobileNetV2, with its focus on economy and speed, is particularly well-suited for embedded and mobile devices. Its operation is based on two basic ideas: depth-wise separable convolutions and inverted residuals. The concept first employs depth-wise separable convolutions to split the convolution process into two stages: a depth-based convolution that applies a single filter to each input channel individually, and the point-wise convolution that combines these filtered outputs. This separation significantly reduces the computational cost as compared to traditional convolutions.

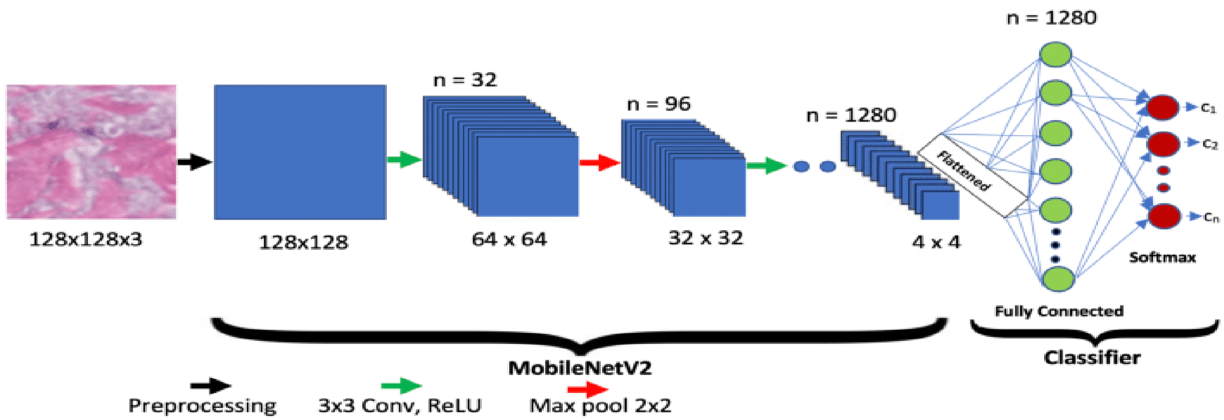


Fig 3.10: The architectural model of MobileNetV2

3.3 Project Plan

Table 3.1: Total project plan & time estimate

Sl. No.	Next Task	Estimate completion time (MM-YY)
1	Data collection	07-24
2	Complete very well of data preprocessing.	08-24
3	Choosing deep learning models deploying.	09-24
4	Reaching the greatest accuracy levels of more than 90%, web application design	10-24
5	Report writing	11-24

3.4 Task Allocation

Tasks	**Weeks**																	
	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Data Collection																		
Data Preprocessing																		
Deploy Models																		
Results																		
Develop web application for prediction																		

Estimated Work Period	
Actual Work Period	

3.5 Summary

The data gathering must come from the completion of all further operations in order to guarantee correctness. I separated the task into its most crucial parts to make it easier to complete. To make sure that my job is completed correctly, I must adhere to these rules.

- Groups of datasets.
- Steps performed in advance of picture processing
- Diagnosing three blood cancer class photographs.
- The many algorithms that are in use.

I openly assembled all of my datasets using programs like Kaggle. I chose a dataset that has information on the various forms of blood cancer. After that, I could go to work on getting the data ready. In this case, I eliminated any superfluous details in my data, such as noise, incorrect images, incorrectly scaled images, etc. Furthermore, I conduct the lengthy durations for training, testing, and validation using datasets.

CHAPTER 4

Implementation and Results

4.1 Environment setup

Blood cancer-based algorithms have predicted a lot of discoveries. Consequently, I used a variety of strategies. Before determining the best plan. Attempting to raise the standard of my work, I experimented with several methods. used the freely available datasets on Kaggle for three blood pictures while getting advice. The three sets of images show various phases of plant growth. I took advantage of the dictionaries, content categorization techniques, and Python tools that were already available. In order to identify the proper cancer cell classifications, this project uses Python DL modules for deep learning to classify images of blood. A number of significant metrics were employed to evaluate the model's performance:

Loss: Both the training and validation phases took into account categorical cross-entropy. The model was becoming more better at identifying therapeutic blood, as seen by a decrease in loss.

Accuracy: The ratio of the total number of response samples to all the samples was used to compute the accuracy evaluating classification, which served as a general indicator of the model's efficacy.

Confusion matrix: The occurrences of each class, including false positives, false negatives, true positives, and true negatives, were identified and predicted by analyzing the model's behavior using the confusion matrix. This matrix showed a number of possible research directions as well as potential obstacles in the search for blood cancer.

4.2 Performance/ Comparative Analysis

I'll employ DL algorithms in this project to predict blood cancer. Every phrase in every subject of study should be given a significant amount of weight in the categorizing process. The datasets were also divided into successive classes using the DL models. One of the most crucial components of any study is the data. I used a range of DL technique approaches and accuracy ratings to help us reach our objective. For this project, I employed five distinct algorithms in all. I then demonstrated, as I mentioned before, that every algorithm was correct. By applying both of these strategies, I was able to get the highest accuracy of 100% in all model tests, which comprise the following three types of blood cancer.

Precision: The degree of precision in which the algorithm generates accurate forecasts is a commonly used metric to assess the efficacy of the model. To calculate efficiency, one can multiply the total number of true positives by the total number of correct forecasts.

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_(i)}$$

Recall: Regardless of all pertinent instances, retrieval is the proportion of appropriate cases that were eventually located and retrieved. When a method has a high recall rate, it is considered to have produced the most relevant findings.

$$\text{recall} = \frac{TP}{TP+FN} \quad \text{\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_(ii)}$$

F1-Score: The accuracy and recall of a test are used to determine its validity. Accuracy and recall work best together.

$$\text{F1} = \frac{2 \times precision \times recall}{precision + recall} \quad \text{\_\_\_\_\_\_\_\_\_\_\_\_\_(iii)}$$

Accuracy: The relationship between an expected value and an acceptable cost is known as reliability.

$$\text{accuracy} = \frac{TP+TN}{TP + FN + TN + FP}$$

4.3 Result & Discussion

My classification processes affected the results I obtained. I have identified blood pictures' exact locations using five different deep learning methods. I utilized deep learning techniques (InceptionResNetV2, MobileNetV2, InceptionV3, VGG16, and VGG19) that have been capable of accurately identifying blood cancer over 20 historical periods. The same dataset, which included both my personal online dataset and data that was publicly available through Kaggle sources, was used by all the models. Using the pre-made Mat-lab libraries, I evaluated the algorithms' soundness after completing the dataset process.

Table 4.1: Accuracy

Models	Accuracy Score
InceptionResNetV2	100%
MobileNetV2	100%
InceptionV3	100%
VGG16	100%
VGG19	100%

The efficaciousness for many models is presented in the next section. PyCharm and CoLab, two open-source tools, were utilized throughout the procedure. There were five models in total: VGG16, VGG19, InceptionV3, MobileNetV2, and InceptionResNetV2. The best-performing models were all DL models, which performed with 100% accuracy.

```
                              precision    recall  f1-score   support

Acute Lymphoblastic Leukemia       1.00      1.00      1.00       194
      Acute Myeloid Leukemia       1.00      1.00      1.00       219
Normal Blood Cell Neutrophil       1.00      1.00      1.00       187

                    accuracy                           1.00       600
                   macro avg       1.00      1.00      1.00       600
                weighted avg       1.00      1.00      1.00       600
```

Fig 4.1: Classification Report

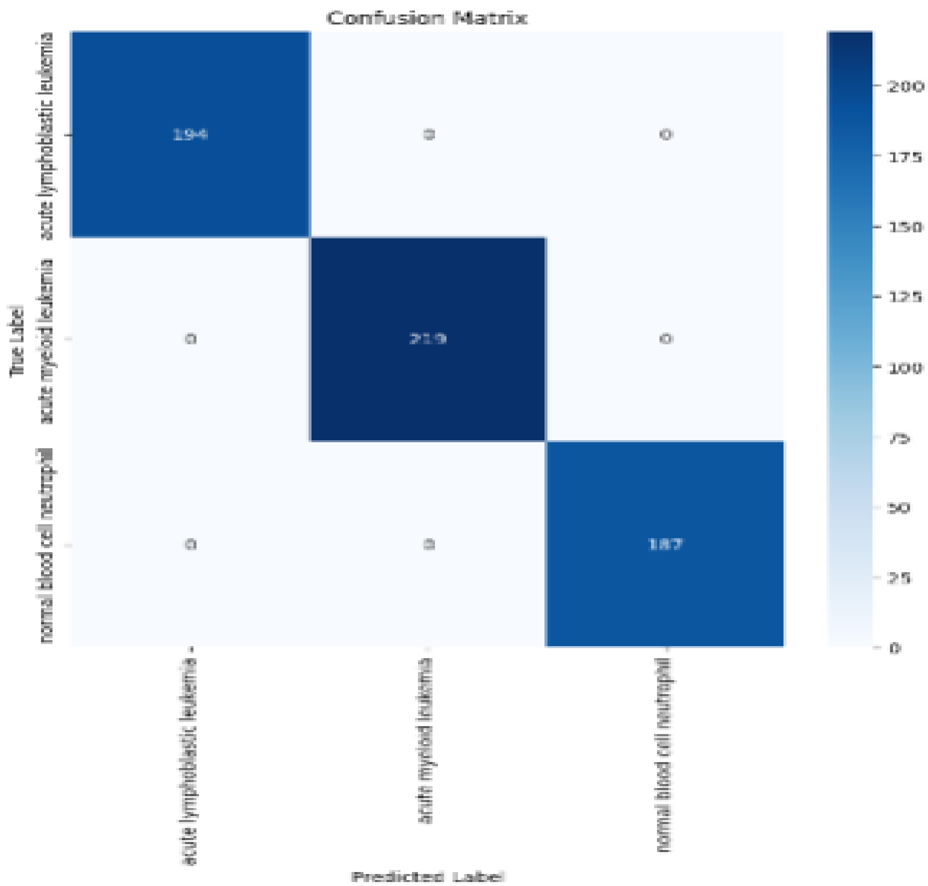


Fig 4.2: Confusion matrix

For the DL models, only the whole classification result is displayed in order to obtain optimal accuracy. The procedures needed to use the CNN approach model to develop a workable classifier for a specific kind of blood cancer are depicted in Fig. 4.3 below.

```
# Example usage
local_image_path = "/content/lympho.jpg"  # Replace with your local image path
img_array = load_local_image(local_image_path)
class_labels = ["Acute Lymphoblastic Leukemia", "Acute Myeloid Leukemia", "Normal Blood Cell Neutrophil"]

predicted_class = predict_class(model, img_array, class_labels)
print(f"Predicted Blood Class: {predicted_class}")
```

```
WARNING:absl:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics
1/1 ━━━━━━━━━━━━━━━━━━━━ 2s 2s/step
Raw predictions: [[9.9902642e-01 3.6490607e-04 6.0863112e-04]]
Predicted class index: 0
Predicted Blood Class: Acute Lymphoblastic Leukemia
```

```
# Example usage
local_image_path = "/content/Myeloid.jpg"  # Replace with your local image path
img_array = load_local_image(local_image_path)
class_labels = ["Acute Lymphoblastic Leukemia", "Acute Myeloid Leukemia", "Normal Blood Cell Neutrophil"]

predicted_class = predict_class(model, img_array, class_labels)
print(f"Predicted Blood Class: {predicted_class}")
```

```
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 75ms/step
Raw predictions: [[8.6082233e-05 9.9988651e-01 2.7381813e-05]]
Predicted class index: 1
Predicted Blood Class: Acute Myeloid Leukemia
```

Fig 4.3: Blood cancer detection

The train and validation accuracy and loss at an epoch of 20 for the Mobilenetv2 models are displayed in Figures 4.7 and 4.8 below.

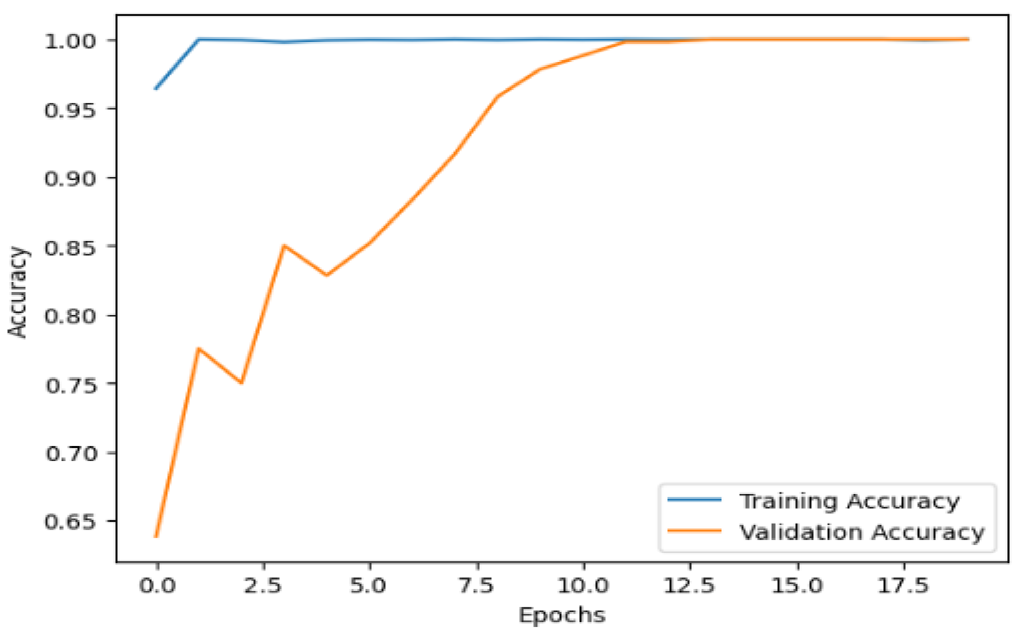


Fig 4.4: Validation and Training Accuracy Curve of MobilenetV2.

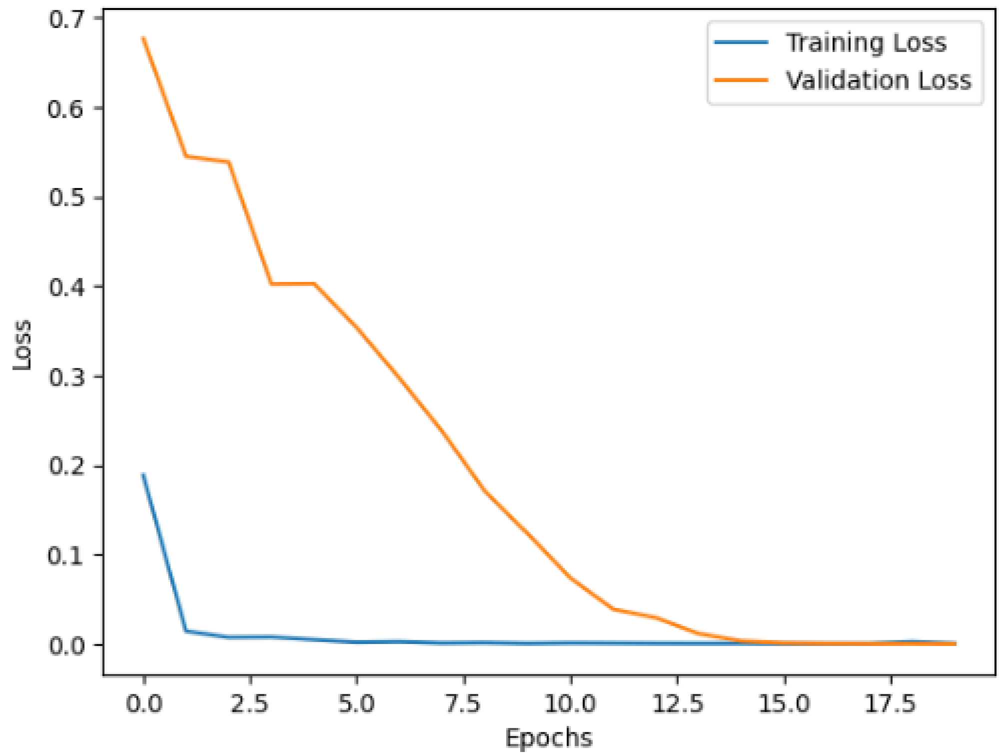


Fig 4.4: Validation and training Loss Curve of MobilenetV2.

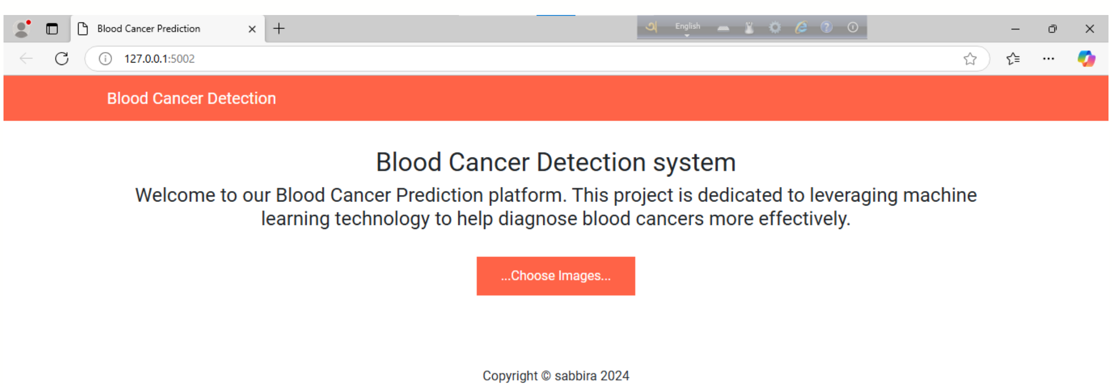


Fig 4.5: Web application prototype for blood cancer detection.

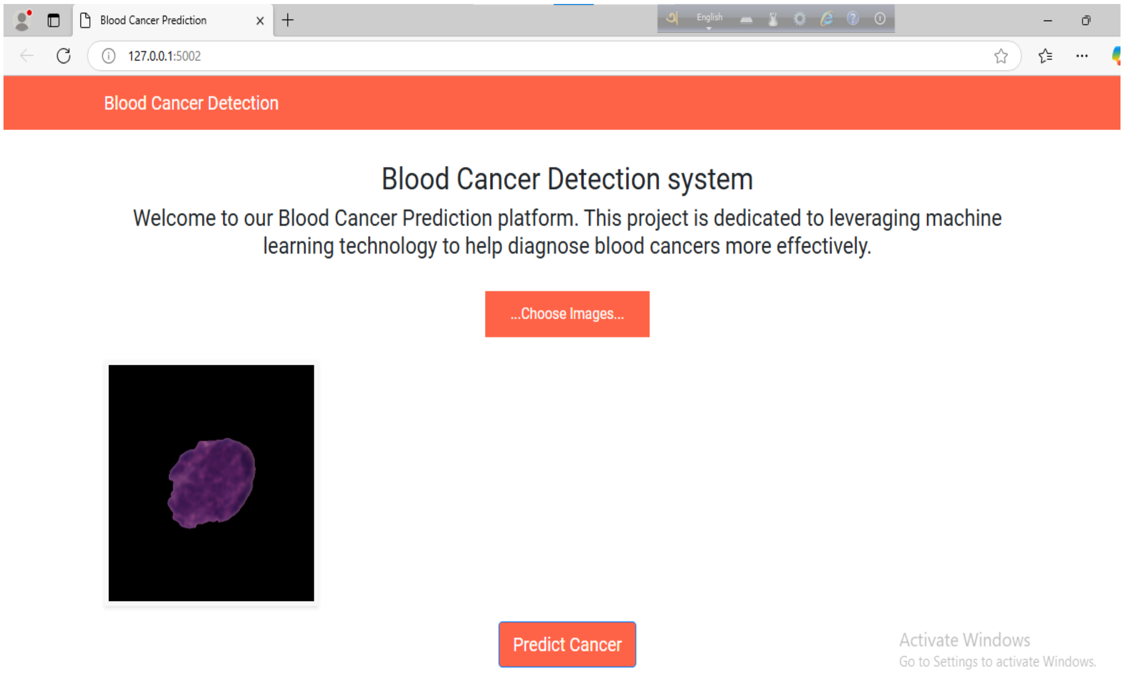


Fig 4.6: After choosing an image from dataset.

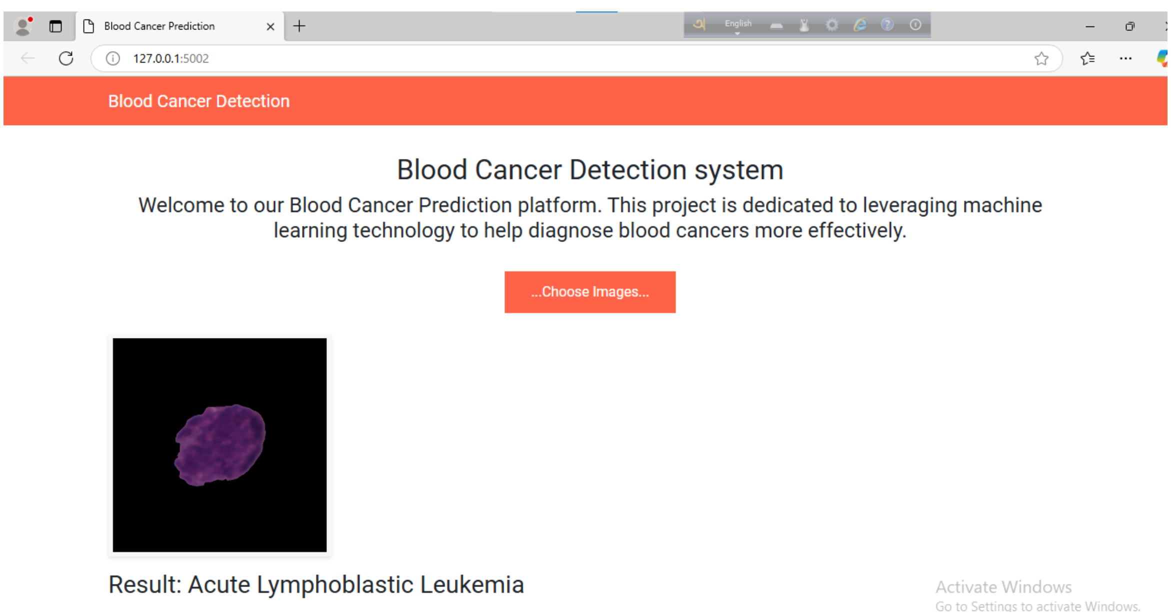


Fig 4.7: Predict Acute Lymphoblastic Leukemia.

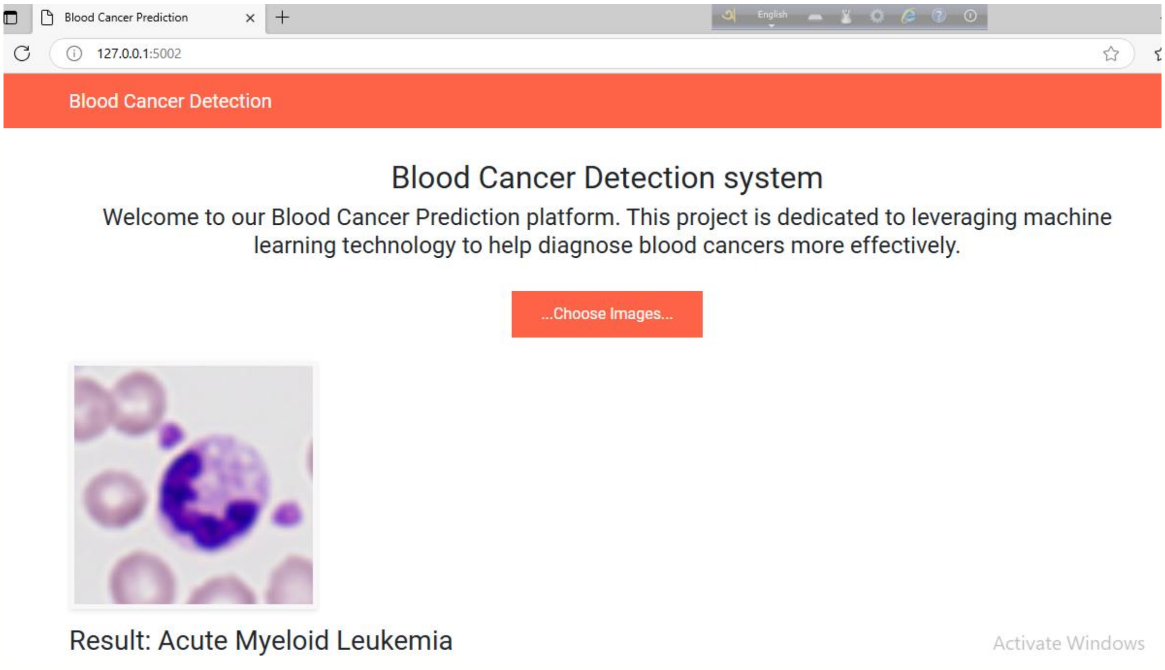


Fig 4.8: Predict Acute Myeloid Leukemia

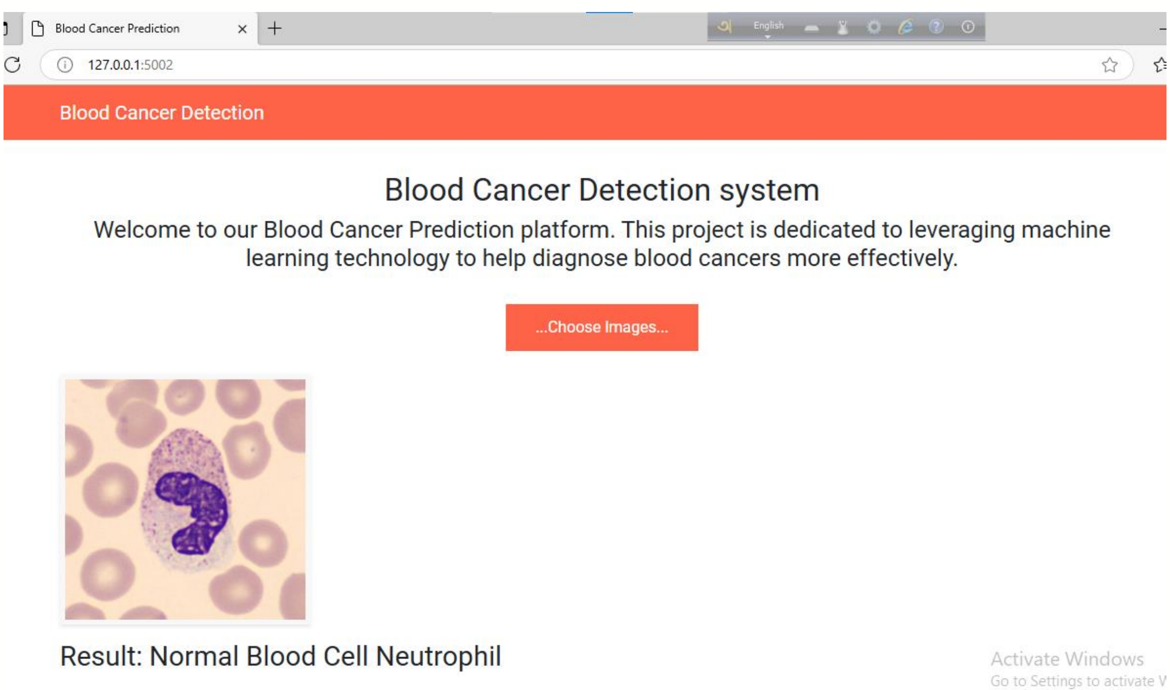


Fig 4.9: Predict Normal Blood Cell Neutrophil.

4.4 Summary

In this section, an image classification approach for illnesses is used to depict the progression of a blood malignancy. The method of developing the model involved gathering suggestions depending on the subject's accuracy. The part that follows presents and discusses the investigation's results. In the phase before, we spoke about the dataset and dataset management strategies. This chapter covers the outcomes of the computations that are performed using the prepared information. The findings measure down to identify which calculation yields the best accurate result. These calculations—VGG16, InceptionV3, vgg19, InceptionResnetv2, and MobileNetv2—are employed.

CHAPTER 5

Engineering standards and Design Challenges

5.1 Compliance with the standards

Following established medical, ethical, and technological guidelines is necessary to guarantee patient safety, data privacy, and model efficacy when it comes to blood cancer diagnosis using deep learning. This entails abiding by regulations set forth by agencies like the FDA or EMA for AI-based diagnostic tools, making sure models are verified through clinical trials, and respecting healthcare standards like HIPAA for data protection. Deep learning models must also be easily integrated into current healthcare workflows and be clear and explainable in order to win clinician trust. These requirements must be fulfilled for AI models to be used in clinical settings for the safe and efficient diagnosis of blood cancer.

5.2 Impact on Society, Environment and Sustainability

5.2.1 Impact on Life

By combining online data from several sources, deep learning techniques are applied to detect cancer cells and increase the accuracy of blood cancer prediction.

- **Enhance the Medical Sector**: The capacity to identify blood cancer cells and advance the field with better outcomes is one of the most powerful potential consequences of deep learning approaches. Early cancer identification will help doctors, patients, and medical professionals discover a remedy, which will contribute to the quick growth of medical knowledge. This might eventually lead to better cancer treatments and more medical output.
- **Economic benefits:** Using DL techniques to identify blood cancer cells may have a positive economic impact on society. For instance, it may help patients if the medical industry grows higher. Remote patient monitoring made possible by DL-powered technology allows for regular status updates without the need for in-person visits. Early treatment may prevent the need for more involved and

costly therapies, which can save money for patients and healthcare systems alike. Additionally, it could lead to more effective medical care and greater financial results. This type of problem will create a critical market for cancer prognostic patients.

- **Social benefits:** Using DL approaches to identify blood cancer advantages. If a cancer patient has a history of blood cancer, their body's immunity will be significantly impacted by blood damage and a reduction in blood production. For instance, cancer patients could benefit from it.

In general, society will be greatly impacted by the technology used to identify leukemia cells utilizing deep learning and machine learning. Reducing blood cancer cells and the medical sector, healthcare sector, all of which would enhance the wellbeing and standard of care for patients both locally and internationally.

5.2.2 Impact on society & Environment

Separating leukemia from cancer cell data using deep learning and machine learning techniques will have a major negative impact on the environment. However, there is a possibility that the development and application of these strategies might unintentionally harm the environment by increasing source consumption. The following are some potential environmental problems that these approaches might cause:

- **Energy consumption**: Using DL to identify blood cancer requires different amounts of energy depending on the particular techniques employed, the dataset's size, the processing power at hand, and the hardware. It has to do with the kind of power employed in calculations. Energy from renewable resources could have less of an impact on the environment than energy from non-renewable sources. Medical outcomes can be enhanced by deep learning for blood cancer detection, but energy implications also need to be considered. There are efforts on to develop more environmentally friendly algorithms and technologies in an effort to decrease the environmental effect of deep learning applications in medical care.

- **Data storage**: The requirement to store this data may require the usage of resources, such as energy and materials, which might be harmful to the environment. It is crucial to take into account the possible environmental repercussions of preserving knowledge as well as possible resource-saving measures in order to mitigate these effects.
- **Transportation:** If distributed deep learning gathers more data and the techniques being employed are accessible in many locations worldwide.

It is improbable that the environment will be directly impacted by the use of deep learning CNN techniques to detect blood cancer utilizing information from many online sources. However, it's crucial to take into account any potential environmental effects of developing and utilizing such technology, and to take all necessary precautions to reduce them. This might mean making use of cost-effective tools and techniques, protecting data securely with the least amount of resources needed, and, if appropriate, lowering transportation-related pollution to cut down on emissions and resource consumption.

5.2.3 Ethical Aspects

DL techniques, as well as online data collecting, can be used to identify blood cancer. A list of some of the most important ethical considerations is as follows:

- **Consent that is informed**: Patients ought to be educated about how machine learning algorithms may be used to diagnose them, the implications, and their opportunity to offer informed consent. They also need to be made aware of any potential risks, benefits, and remedies.
- **Bias:** Preexisting prejudices in healthcare may be inadvertently reinforced or amplified by ML models. Accountable data must be used in training, and no one demographic may be unjustly favored or penalized by the framework.

- **Allocating Efforts:** The application of machine learning in healthcare shouldn't take funds away from other vital areas of patient care or exacerbate treatment access gaps that currently exist.
- **Safety and Privacy**: It is important to take the utmost care while handling patient information, especially in the healthcare sector. Ethical concerns include safe data storage, suitable encoding, and strict access control in order to protect patient confidentiality.
- **Simplicity and Explainability**: Machine learning models, especially for those who don't, can be hard to understand and intricate. Maintaining the highest ethical standards requires the model to have a well-defined decision-making process and be able to justify its forecasts.

All things considered, the application of DL in the diagnosis of blood cancer may greatly enhance the course of individual patient care; however, in order to ensure that this technology is applied sensibly, equitably, and in a manner that prioritizes the best interests of patients, it must be managed with great care and due diligence.

5.2.4 Sustainability Plan

Sustainability-focused deep learning blood disease detection necessitates taking social, economic, and environmental factors into account. For this project, the following is a sustainable development plan:

- **Data security and privacy**: Ensure that private patient information is sent and kept secure in order to respect legal and ethical obligations while also safeguarding patient privacy.
- **Precision and dependability**: Ensuring The accuracy and reliability of the ml engines determine the strategies' long-term viability. involves closely reviewing and validating the processes to ensure their accuracy as well as routinely evaluating and enhancing the models as needed.
- **Accountability:** The assignments and duties If they are to be effective, those involved in the development and use of DL techniques need to be well-defined.

This might entail outlining the precise conditions for the methods' moral application as well as the responsibilities and duties attached to them that go along with making sure they are applied both safely and ethically.

- **Algorithm efficiency**: Reduce the amount of computational power needed for training and inference by utilizing techniques like knowledge purification, quantization, and modeling trimming. Set up a schedule for model updating and retraining to ensure that they stay accurate and useful over time.

By incorporating these sustainability factors into the usage of DL models for blood cancer detection, one may improve patient satisfaction while lowering the environmental impact and contributing to the creation of a better healthcare system.

5.3 Thesis Management and Financial Analysis

Our aim is connected to the deep learning project we are working on. This section's portion on our project's financial analysis is required. The expected expenditures for each component of the deep learning project are summarized in Table 3.1 below.

Table 3.2: Estimated Cost for Blood cancer prediction

SN	Components	Estimated Cost (BDT)
01.	Visiting Stakeholders	500-1000
02.	Software and Tools	1500-2000
03.	Data Collection and Processing	500-1000
04.	Documentation and Report Writing	500-1000

05.	Contingency (10% of total)	1500-2000
	Total Estimated Cost	4,500-7,500

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

Create a category-based mapping to solve this area's problems. Provide subsections for each mapping to support your claims (see Table 5.1).

Table 5.1: Mapping with complex problem solving.

EP1 Dept. of Knowledge	EP2 Range of Con-flicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent of Stake-holder Involvement	EP7 Interdependence
√	√	√	√	√	√	√

Mapping with Knowledge Profile for EP1

This table 5.2) is designed to map the EP1 to the Knowledge Profile.

Table 5.2: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
√	√	√	√	√

5.4.2 Engineering Activities

In this part, provide a mapping of engineering activities. Provide subsections for each mapping to support your claims (see Table 5.3).

Table 5.3: Mapping add subsections to put rationale.

Engineering Activity	Rationale
1. Data Collection	The dataset was gathered via the internet resource Kaggle. I made a dataset of 6000 images. Based on the kind of sickness each class in the collection reflects, it is separated into three groups with 2000 photographs each. 3 types of blood sample like: Neutrophil-containing normal blood cells, acute myeloid leukemia, and acute lymphoblastic leukemia are among the classifications.
2. Data Preprocessing	The results expanded and narrowed as each class was assessed. I had to alter the size and add information for it to work. I made just the biggest and most socially acceptable changes since I was concerned about overfitting.
3. Model Development	Once you've made your decision, train and assess the selected model using the provided data to increase accuracy. DL uses a wide variety of models. Using my technology, I experimented with many versions of the concept to determine the best configuration for precise data calculation.
4. Model Evaluation and Testing	A description of each outcome is given in this section. Even after testing and training, these strategies left us lacking adequate reliability for the next two courses. They created a technique for classifying different medicinal plants and created graphics for f1 measurement, recall, efficiency, and confusion matrix.
5. Deployment and Integration	Real-time predictions are made possible by integrating the model into an intuitive platform (such as a web or mobile app). Access control and encryption are two essential security techniques for safeguarding private health information.

6. Continuous Monitoring and Maintenance	Continuous observation guarantees the model's accuracy over time. Maintaining the model's relevance and conformity to changing standards is facilitated by frequent updates based on fresh data and performance feedback.

5.5 Summary

This section offered a summary about the study, Engineering Standards and Design Challenges, compliance with standards, financial analysis, engineering problem, information on the upcoming events, and a question regarding the results. This chapter offers an authorized sample to show whether or not the supplied report satisfies all requirements. Impacts on the surroundings and the group as a whole: This chapter on and how they may not meet the needs of future workers pursuing similar goals.

Table 5.4: Mapping with complex engineering activities.

EA1 Range of re-sources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
√	√	√	√	√

CHAPTER 6

Conclusion and Future Work

6.1 Summary

Based on the study, the research findings and methodologies I employed are excellent. It is my belief and hope that this study's findings will stimulate further research in this area. In order to categorize blood cancer and identify it through the use of a deep learning model, five distinct methods have been employed. For my specific dataset. In this instance, a dataset I put together from several web sources served as the primary source of data. I will be able. I'm going to have a ton of ideas for growing my job because of this study. While working, I came across a few mistakes. I found further directions that this investigation may go. It will enable us to address any defects or other issues that come up while I work on this project. In addition, I'm considering how to better address the issues this study presents by including algorithms into my future work. This research will allow me to learn more about our subject of study. I think it will help with blood cancer research and innovative tech solutions that let us help the medical sector. Based on this study, I hope to present a new blood cancer detection approach that will help the medical community use DL CNN methods to categorize cancer. Three categories of blood cancer photos are used: "Acute Myeloid Leukemia," "Acute Lymphoblastic Leukemia," and "Normal Blood Cell Neutrophil." Six thousand pictures were utilized in this investigation. During the process, we presented and contrasted a number of deep learning models, including VGG16 & VGG19, Inception V3, and Mobile Net V2. The assessments conducted on each of the five models demonstrated an average accuracy of 100% on the dataset. Develop a technique for classifying blood cancer based on the model.

6.2 Limitations

DL models can examine large databases to detect blood cancer early, when treatment is more likely to be successful. This might lead to better patient outcomes and higher survival rates providing targeted therapy to individuals who most need it. Early diagnosis and tailored treatment may be able to reduce overall healthcare costs by removing expensive and time-consuming therapies in the later stages of the illness. We tried our best, but we were unable to gather further types of data during this time. So, this is a limitation. We just took pictures of the leaves' front side; we should have also taken a picture of their reverse side. This is an extra limitation. Furthermore, even though we have used Deep CNN approaches, we may still investigate other Machine Learning models.

6.3 Future Works

More research in this area may go in a lot of different ways, particularly in the context of my own job. I was full of ideas about how we might improve our job. As previously said, I have discovered a few mistakes as well, and these mistakes offer opportunities to improve this study. By implementing this work and enhancing prediction results with a higher likelihood of capturing high accuracy, I'll attempt to correct this issue. I intend to correct any mistakes that occur. I have other objectives in mind. There is room for improvement in this section of the prediction. I will include components like cancer detection types using the CNN model and other factors into the algorithm I use to develop the software that users may employ to forecast blood cancer in order to help the user and make the most out of this. I think I can use this kind of work to progress medical technology and increase its important impact on healthcare. I could help the people be diagnosed with cancer early.

Reference

[1] C. Matek, S. Schwarz, C. Spiekermann, and K. Marr," Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks." *Nature Machine Intelligence*, *1*(11), 2019, pp.538-544.

[2] R. Duggal, A. Gupta, R. Gupta and P. Mallick, "SD-layer: stain deconvolutional layer for CNNs in medical microscopic imaging." In *Medical Image Computing and Computer Assisted Intervention− MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20* ,2017, (pp. 435-443). Springer International Publishing.

[3] J. Prellberg, and O. Kramer, "Acute lymphoblastic leukemia classification from microscopic images using convolutional neural networks." In *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging: Select Proceedings,* 2019, (pp. 53-61). Springer Singapore.

[4] V. Rupapara, F. Rustam, W. Aljedaani, H.F. Shahzad, E. Lee, and I. Ashraf, "Blood cancer prediction using leukemia microarray gene data and hybrid logistic vector trees model." *Scientific reports*, *12*(1), p.1000,2022.

[5] S. Shah, W. Nawaz, B. Jalil, and H.A. Khan, "Classification of normal and leukemic blast cells in B- ALL cancer using a combination of convolutional and recurrent neural networks." In *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging: Select Proceedings*, 2019, (pp. 23-31). Springer Singapore.

[6] H. Akash, H. Naveena, M. Rakesh, P. Yallalinga and M.Suhmitha, "Blood Cancer Detection Using Machine Learning." Vol. 9, Issue 6, 2021. DOI: 10.17148/IARJSET.2022.9639.

[7] K.P. Jayavikash, P. Gopinath, U. Nandheeswaran, R. Sivabalan, and T.D. Bharathi, "Detection of Leukemia Using Machine Learning Algorithms." In *Journal of Physics:Conference Series*,2021, May(Vol. 1916, No. 1, p. 012220). IOP Publishing.

[8] M. Iswarya, “Detection of Leukemia using Machine Learning.” In *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* ,2022, May, (pp. 466-470). IEEE.

[9] F.Mohd, N.M.M.Noor, Z.A. Bakar, and Z. A. Rajion,“Analysis of oral cancer prediction using features selection with machine learning.” In The 7th International Conference on Information Technology (ICIT)(2015).

[10] Y. Xiao, J. Wu, Z. Lin, & X. Zhao, “A deep learning-based multi-model ensemble method for cancer Prediction.” Comput. Methods Programs Biomed. 153, 1–9 (2018).

[11] O. Gal, N. Auslander, Y. Fan, & D. Meerzaman, “Predicting complete remission of acute myeloid leukemia:” Machine learning applied to gene expression.” Cancer Inform. 18, 1176935119835544 (2019).

[12] M. Ghaderzadeh, F. Asadi, A. Hosseini, D. Bashash, H. Abolghasemi, and A. Roshanpour, “Machine learning in detection and classification of leukemia using smear blood images: a systematic review.” *Scientific Programming*, *2021*, pp.1-14.

[13] R.B. Hegde, K. Prasad, H. Hebbar, B.M.K Singh, and I. Sandhya, “Automated decision support system for detection of leukemia from p eripheral blood smear images. *Journal of digital imaging*, *33*, 2020, pp.361-374.

[14] D. Kumar, N. Jain, A. Khurana, S. Mittal, S.C. Satapathy, R. Senkerik, and J.D. Hemanth, “Automatic detection of white blood cancer from bone marrow microscopic images using convolutional neural networks. *IEEE Access*, *8*, 2020, pp.142521-142531.

[15] L.S. Chen, A.M.F. Yen, S.Y.H Chiu, C.S. Liao, and H.H. Chen, “Baseline faecal occult blood population concentration as a predictor of incident colorectal neoplasia: longitudinal follow-up of a Taiwanese population-based colorectal cancer screening cohort.” *The lancet oncology*, *12*(6),2011,pp.551-558.

[16] S. Jagadeesh, E. Nagabhooshanam, and S. Venkatachalam, "Image processing based approach to cancer cell prediction in blood samples." *International Journal of Technology and Engineering Sciences*, *1*(1), 2013, pp.1-10.

[17] E. Racila, D. Euhus, A.J. Weiss, C. Rao, J. McConnell, L.W. Terstappen, and J.W. Uhr, 1998. "Detection and characterization of carcinoma cells in the blood." *Proceedings of the National Academy Of sciences*, *95*(8), pp.4589-4594.

[18] C.R.V. Dotimas , J.R.Z. Erne, A.J.H. Jamig, L.B. Tarronas, and A.T. Magpantay, " PREDICTION OF POSSIBLE LEUKEMIA IN MICROSCOPIC BLOOD CELL IMAGES: A MACHINE LEARNING AND IMAGE PROCESSING APPROACH."

[19] G.D. Poore, E. Kopylova, Q. Zhu, C. Carpenter, S. Fraraccio, S. Wandro, T. Kosciolek, S.Janssen, J. Metcalf, S.J. Song, and J. Kanbar , "Microbiome analyses of blood and tissues suggest cancer diagnostic approach." *Nature*, *579*(7800), 2020, pp.567-574.

BLOOD CANCER DIAGNOSIS USING DEEP LEARNING: ENHANCING ACCURACY IN LEUKEMIA DETECTION

ORIGINALITY REPORT

13%	6%	4%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Southern Mississippi Student Paper	4%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
3	Submitted to Daffodil International University Student Paper	1%
4	"ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging", Springer Science and Business Media LLC, 2019 Publication	1%
5	digital.library.unt.edu Internet Source	<1%
6	Submitted to Pennsylvania State System of Higher Education Student Paper	<1%
7	"Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017", Springer Nature, 2017 Publication	<1%